Security Considerations for Computational and Data Grids

William E. Johnston[a], Keith Jackson[b], and Sumi Talwar[c]

Abstract

Large-scale science and engineering is frequently done through the interaction of collaborating groups, heterogeneous computing resources, information systems, and instruments, all of which are geographically and organizationally dispersed.

"Grids" provide the services needed for building dynamically constructed problem solving environments using geographically and organizationally dispersed high performance computing and data handling resources. The overall motivation for "Grids" is to enable the routine interactions of these resources to enhance this type of large-scale science and engineering, and thus substantially increase the computing and data handling capabilities available to science and engineering projects.

Grids present a very different environment from that of traditional computing and data handling. We are moving from the remote login and data transfer approach to dynamically building application specific systems that are based on many widely distributed components that are owned and operated by many different institutions.

However, even if this environment works in every other way, it will not be viable if it is constantly disrupted by hackers and their kin. Distributed applications are potentially more vulnerable than conventional scientific problem solving environments because there are substantially more targets to attack in order to impact a single application.

Much of the overall security of Grids is inherited from the security of the underlying systems. There are, however, some security considerations at the Grid level that are independent of the underlying systems, and we focus on this later aspect.

[a] Lawrence Berkeley National Lab and NASA Ames Research Center, wejohnston@lbl.gov (corresponding author)
[b] Lawrence Berkeley National Lab, krjackson@lbl.gov
[c] NASA Ames Research Center, sumit.talwar@amti.com

1 The Grid Environment

Large-scale science and engineering is frequently done through the interaction of people, heterogeneous computing resources, information systems, and instruments, all of which are geographically and organizationally dispersed.

The overall motivation for "Grids" [1] is to enable the routine interactions of these resources to enhance this type of large-scale science and engineering.

Functionally, Grids are tools, middleware, and services that:

- provide a uniform look and feel to a wide variety of distributed computing and data resources
- support construction, management, and use of widely distributed application systems
- facilitate human collaboration and remote access to, and operation of, scientific and engineering instrumentation systems
- provide for managing and securing this computing and data infrastructure

This is accomplished through a set of *uniform software services* (the Common Grid Services) and that may be summarized as

• information services	• resource specification and request
• resource co-scheduling	• data access
• authentication and authorization	• security services
• auditing	• monitoring
• global event services	• global queuing
• data cataloguing	• resource brokering
• collaboration and remote instrument services	• data location management
• communication services	• fault management

See Figure 1.

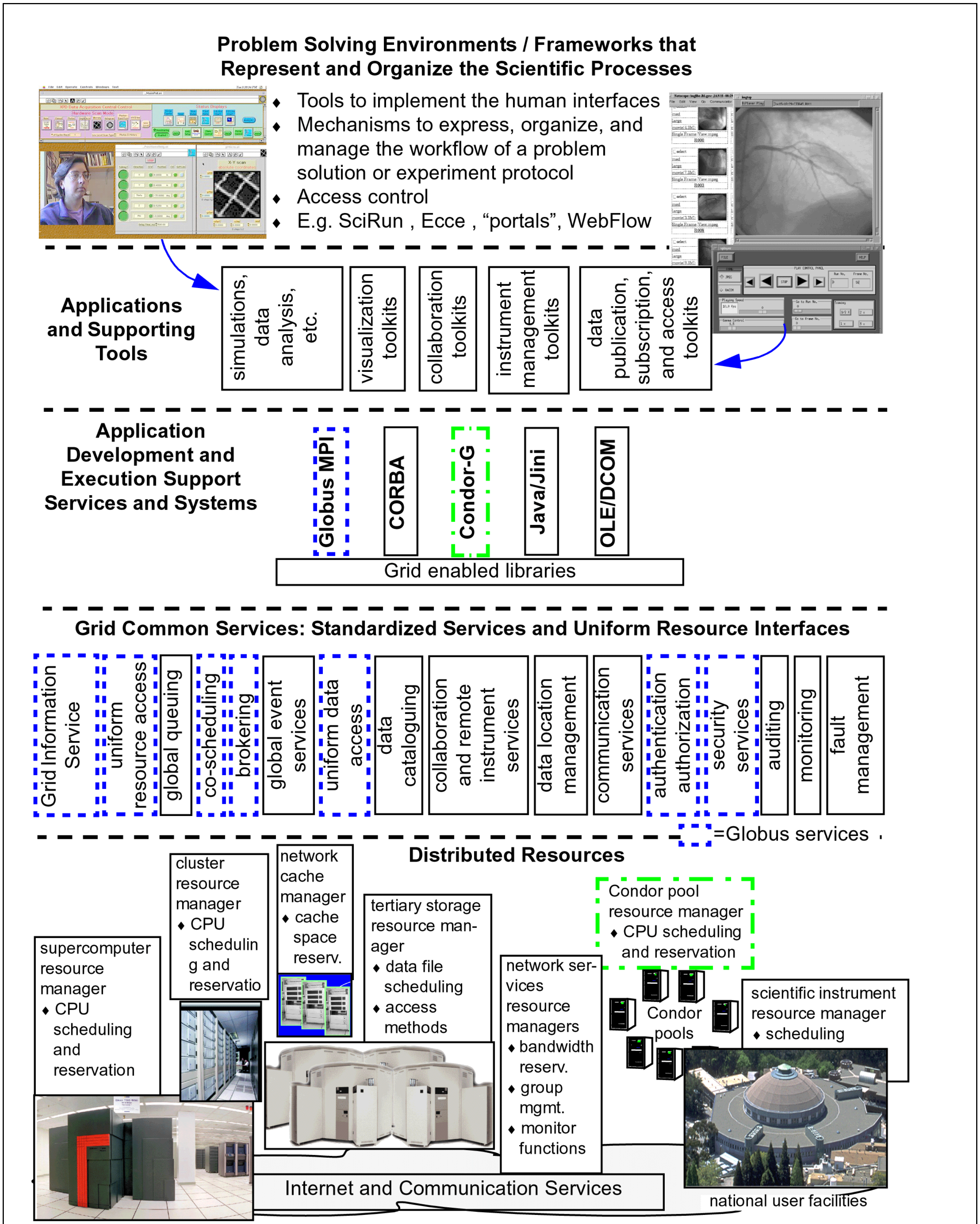

Figure 1. A Grid Architecture

1.1 Applications

The overall motivation for the current large-scale (multi-institutional) Grid projects is to enable the resource interactions that facilitate large-scale science and engineering such as aerospace systems design, high energy physics data analysis, climatology, large-scale remote instrument operation, etc.

The vision for computing, data, and instrument Grids is that they will provide significant new capabilities to scientists and engineers by facilitating *routine* construction of information based problem solving environments that are built on-demand from large pools of resources. That is, Grids will routinely – and easily, from the user's point of view – facilitate applications such as:

- coupled, multidisciplinary simulations too large for single computing systems (e.g., multi-component turbomachine simulation – see [2])
- management of very large parameter space studies where thousands of low fidelity simulations explore, e.g., the aerodynamics of the next generation space shuttle in its many operating regimes (from Mach 27 at entry into the atmosphere to landing)
- use of widely distributed, federated data archives (e.g., simultaneous access to metrological, topological, aircraft performance, and flight path scheduling databases supporting a National Air Transportation Simulation system)
- coupling large-scale computing and data systems to scientific and engineering instruments so that complex real-time data analysis results can be used by the experimentalist in ways that allow direct interaction with the experiment (e.g. Cosmology data analysis involving telescope and satellite interaction, and coupling to simulations)
- single computational problems too large for any single system (e.g. extremely high resolution rotocraft aerodynamic calculations)

1.2 Characteristics of the Environment

Consider two examples of Grid-like applications that illustrate the environment.

1) Real-time digital libraries for on-line, high data-rate instruments [4] (data intensive computing: use of widely distributed, federated data archives – see Figure 2). These are characterized by:
   - on-line, real-time, high data-rate medical instrument
   - management of large data sets in wide area
   - remote data analysis followed by automatic data cataloguing and archiving
   - remote data users
   - widely distributed, high performance "application-level" cache
   - strict authorization and access control

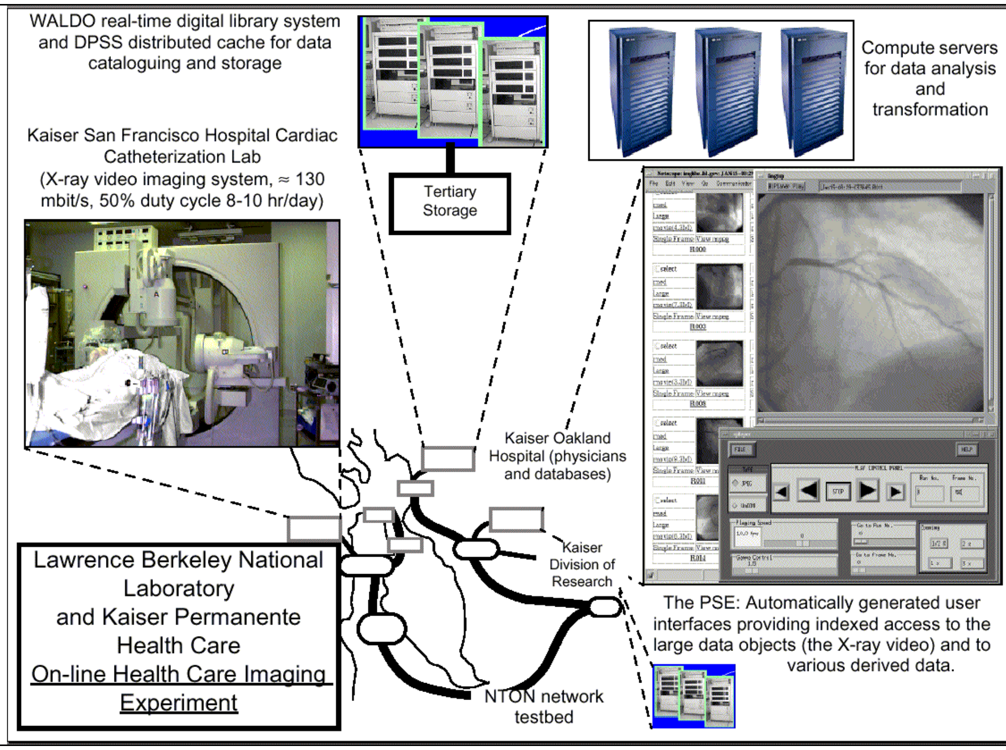

Figure 2. On-Line Medical Imaging System

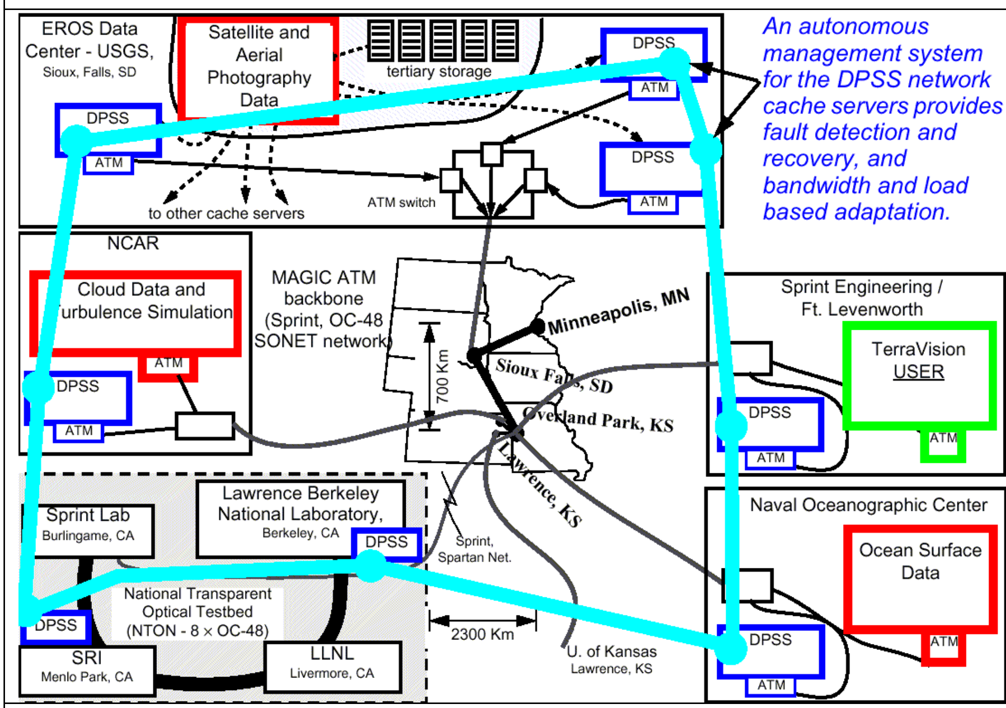

Figure 3. The MAGIC Testbed Distributed Application Environment

2) High data-rate distributed data management and federated access for archived satellite and aerial imagery, digital terrain data, and atmospheric data ([6] and [7] - see Figure 3).). These are characterized by:
   - on-line, real-time access to multiple environmental data sets that are (and always will be) maintained by domain experts at their own sites.
   - on demand, real-time interactive exploration of an operational environment supporting, e.g., military operations and community emergency services
   - aggregation of multiple, widely distributed, multi-discipline data sets

These large-scale science and engineering problems involve many types of applications and data sources that are accessed and shared across many institutions. This implies:

- numerous interconnected servers providing computational simulation and analysis, data access, and functional access to instruments, in semi-open agency/research networks (e.g., ESNet, NREN, Internet-2, etc.)
- many simultaneous collaborators, e.g. at DOE Labs, NASA Centers, other Federal labs, industrial partners, and many universities (esp. DOE/OSC)
- many stakeholders, diverse assets

2 Overall Security Considerations

By "Grid applications system" we mean the collection of software components and the platforms in a Grid environment that are bound together by virtue of utilization by a single application. This binding could be fairly tight in that the components could be continuously connected by communication channels, or it could be quite loose, with components generating asynchronous events, perhaps without the direct knowledge of the receiver (which is, never the less, part of the overall application from the human perspective). These applications may be transient – being built on demand (or when all of the required resources are available) – and may involve a different set of resources (e.g. computing engines) every time they are (re)constituted.

Some overall security considerations in this sort of an environment are:

- Strong authentication to a globally unique identity

  Users are no longer necessarily listed in a single central database at a local site, however positive identification to an entity that can provide human accountability will still be required in most cases.

- Strong and flexible, policy based authorization and access control

  Grid application systems will be composed of resources that are distributed geographically and organizationally. There will be multiple stakeholders who will probably not have a uniform resource use policy. Therefore users will have to be authorized separately, and perhaps with respect to several different attributes, for every resource that is incorporated into a Grid application system. With many remote users of valuable resources, data, code, etc., that are dispersed across many platforms, automatic evaluation and enforcement of access rights is essential.

- Grid applications should not contribute to, nor inherit, vulnerabilities of the underlying/local systems/resources

Grid services must not weaken security of local systems, and a security compromise on one platform that is involved in a Grid application system should not propagate via Grid services to other platforms in the system.

- Grid users/mgrs will not have control over the security policy of resource platforms (e.g. computing systems) in other administrative domains

  It may be necessary to "rate" systems on their security, and provide that rating as a system characteristic that may be used in choosing resources from a candidate pool when constructing the resource base for a distributed application.

- Grid security should be able to be used to enhance local system security

  Grid security services provide, e.g., X.509 certificate based applications for login and file transfer that can be used for general access to systems.

- Security for the Grid Information Service

  Central resource information catalogues must allow for local control of exported information.

3 Assets to Protect

There are different security environments for Grids depending on their use, and on the nature and security of the underlying systems. For example:

1) General science Grids operating in open network environments on systems without a common security model

   General science Grids, of which the DOE/OSC Science Grid is an example, will have many participants: Labs, universities, industrial partners, etc.

2) Mission critical data Grids where code and data is proprietary, and there is a common security model

   Mission critical Grids will involve mission critical data and code, and will probably operate primarily across mission related institutions such as the NASA Centers.

3) Mission critical operations Grids that provide, e.g., operational control, logistical support for an organization or mission

4) Grids doing national security related work

   ASCI DISCOM ….

The security issues for the Grid aspects of all of these scenarios are similar: authentication, authorization and access control, and mitigation of underlying system vulnerabilities. However, the security issues for the underlying computing, storage, and communications will be quite different for the different scenarios.

In this discussion we will only address 1 (general science Grids) and 2 (mission critical data Grids).

The assets associated with Grids are:

- Grid resource use and/or access is valuable

    - computing systems
    - data management and mass storage systems
    - scientific / engineering instruments
    - collaboration services
    - communications systems
  - Intellectual property is potentially valuable and/or proprietary (scientific, commercial, and national interests)
    - source code and data
    - collaborative interaction
- Cross-site trust is valuable
    - access to resources that can be shared is an important Grid capability

4 Threat = Motivated Adversary + Vulnerability

For general science Grids / mission critical data Grids we can characterize the threat environment as motivated adversary plus vulnerability.

Adversary	Motivation - Probability (Sci Grid / Mission critical data Grid)
foreign government agents - national security espionage	low / low (no national security aspects)
foreign government agents - commercial / intellectual espionage	medium / high (scientific data can be a national asset)
cyber-terrorists	low / medium (not operational mission critical)
commercial espionage	medium / medium-high (do have some proprietary data)
scientific espionage	medium-high / low (professional rivalry)
criminals	low / low (little commercial value, few extortion targets)
Byzantines (insiders, e.g. disgruntled employees)	? (wild card)
skilled hackers (recreational)	medium-high / medium-high
script kiddies	high / high

5 Vulnerabilities, Risks, and Consequences

Vulnerability	Risk	Consequence

OS element mis-design, mis-implementation, mis-configuration / mismanagement	unauthorized system access	theft of service, denial of service for whole system
	root compromise and active wiretap	loss of data/control stream confidentiality and integrity
	theft of identity	acquisition of unauthorized privileges
Application mis-design and mis-implementation	(same as above - rootkit.com)	
Communications elements mis-design, mis-implementation, mis-configuration / mis-management	route corruption	denial of service
	active wiretap	loss of data/control stream confidentiality and integrity
Physical access to systems / infrastructure	Passive wiretap	loss of data/control stream confidentiality
	Active wiretap (man-in-the-middle attacks)	loss of data/control stream confidentiality and integrity
Open IP networks	coordinated attacks saturate network defenses	denial of service at many levels, from basic access to server functions
Compromise of identity certificates	theft of identity	unauthorized privileges
		compromise of signature
Social engineering	theft of identity	unauthorized privileges
		compromise of signature

6 Risk Reduction

Note that:

- Grids are heterogeneous environments - many different resources, many different sites, many different policies
- Most Grid services are, by design, "guests" of the local systems (participating in a Grid does not require extensive modification of the local environment, and this includes the local security environment)

In regards to "Threat = Motivated Adversary + Vulnerability," we can't do much about motivation, so we try and reduce vulnerabilities.

Considering five main sources of vulnerabilities

1) the underlying systems
2) Grid services and applications
3) communications infrastructure
4) users

5) diverse security models for local systems and their servers

then these, together with the access authorization needs of some user communities, leads to a set of requirements.

7 High Level Security Requirements for Grids

Note that in the points below, *m* = mission critical requirement.

A-1) Secure identity management supporting remote authentication and single, global identity sign-on w/o clear text passwords
  - basis of authorization for everything from safe, remote access to Grid resources to critical code and data
  - migration to hardware credential tokens will further reduce risk of theft of identity

A-2) Identity proxy / delegation for access to services managed by third parties
  - due to the diversity and number of Grid resources, individual users will normally rely on services / brokers to locate and engage resources on their behalf – this will require proxy/delegation credentials

A-3) Grid services control channel integrity and confidentiality
  - most server weaknesses relate to control operations, or mixed control and data

A-4) Optional data integrity and confidentiality (in transit, in middleware, and in storage^{m})
  - must be end-to-end (e.g. encryption/decryption at the application layer)

A-5) Policy based authorization providing access control for, e.g., individual, group, and role
  - must provide for diverse / multi-organizational stakeholder management of use conditions, accommodate third party user attribute certification, be capable of flexible integration with a wide range of applications and access control gateways (e.g. see "Certificate-based Access Control for Widely Distributed Resources" (Akenti, [13]).

A-6) A set of high-level access and data movement services that are built on the capabilities of -
  - e.g. Secure Shell, secure ftp, encrypted file I/O^{m}, secure remote file systems, etc.

A-7) Cross-site integrity and cyber risk mitigation
  - compromise of system security at one site should not propagate to other sites, even when Grid services are operating in the compromised environment

A-8) Infrastructure assurance
  - must do whatever is possible to prevent denial-of-service attacks

A-9) Auditing and non-repudiation
  - secure auditing with assured identity is required for many operational tasks, including, e.g., accounting

8 Grid Security Services

The overall strategy is vulnerability reduction through universally and readily available security services that address the requirements:

B-1) Cryptographic identity based authentication for all users
- addresses requirements A-1), A-5), and A-9)

B-2) General security service libraries for applications
- addresses requirements A-3), A-4), A-7)

B-3) General authorization libraries and/or security gateways
- addresses requirements A-5), A-7)

B-4) Integrated security services and security enabled utilities
- addresses requirements A-1), A-2), A-3), A-4), A-5), A-6)[a], A-7)

B-5) Long-term key management
- addresses requirement A-4)[b] (data storage[m])

B-6) Secure IP and secure DNS
- addresses requirements A-8)

B-7) Server application coding standards
- addresses requirement A-7)

B-8) Active (e.g. scanning based) enforcement
- addresses requirement A-7)

Notes:

1) Recall that local system/platform security is a separate issue, and that these services are meant to support a Grid security model that is in addition to local security models.
2) The primary difference in security for science Grids as compared with Mission Critical Grids will be
   - in the strength of the security model implemented in the underlying communications, computing, and storage systems
   - in the policy and operational model for the PKI (e.g. [9] - "Public Key Infrastructure Roadmap for the Department of Defense")
3) Most of the basic security services have been built on both PKI ([10]) and Kerberos, however, in the general science Grid environment that involves many geographically and organizationally dispersed institutions, we have focused on PKI.
4) Part of the reason to "compartmentalize" the servers on individual systems (i.e. requiring server and user re-authentication at each new server) is to reduce the impact of a compromised system (platform) in the middle of a chain of servers.

Considering the services in a little more detail:

---

[a] does not provide encrypted file I/O
[b] not currently available
[m] for mission critical Data grids

B-1) Cryptographic identity based authentication

- o single sign-on (mobility and no clear text passwords - e.g. via PKCS-12 [] and ssh [15])
- o remote verification of identity (e.g. via TLS [14])
- o digital signature for data integrity/authenticity and non-repudiation (when combined with secure timestamping)
- o key management for encrypting channels (e.g. via TLS) and storing data
- o identity certification authorities providing third-party assurance of identity (e.g. X.509 Public Key Infrastructure [10])

B-2) General security service API (e.g., IETF Generic Security Service Application Program Interface - "GSS" [11])

- o message integrity and confidentiality

B-3) General authorization API (e.g., IETF Generic Authorization and Access Control ("GAA") [12])

- o several experimental implementations exist
- o LBNL's Akenti [13] provides authorization based on policy stipulated by many independent stakeholders, and a GAA interface for Akenti is being developed

B-4) Globus Security Infrastructure (GSI) [8] integrates the basic security services with delegation (proxy identity, which is essential for brokered access to resources) to provide:

- o secure services/utilities: remote login, copy, ftp, etc.
- o universally and readily available support for secure communication channels and authorization in applications
- o *(does not currently support encrypted file I/O*[m]*)*

B-5) Long-term key management for data files encrypted while in storage

- o (this service is not currently available[m])

B-6) Secure IP and secure DNS

- o IPsec provides identification, integrity, and confidentiality at the network layer
- o secure DNS with host certificates provides for authentication of hosts
- o can protect against some forms of denial of service attacks
- o *this isn't widely available yet*

B-7) Server application software implementation standards will require the use of authentication and authorization on any channel capable of control operations

B-8) Active (e.g. scanning based) enforcement provides one element of site integrity by identifying non-complying application code / servers

9 Grid Security Model

The primary aspects of the security model are:

- o All command and control functions are transported over encrypted channels, after the client/user is authenticated and authorized.
- o This authorized, encrypted command channel compartmentalizes all servers: If multiple servers are involved in a distributed system, then each reauthorizes connections through the use of cryptographic proxies or active re-authentication.

- Apart from confidentiality, pure data channels are not considered vulnerable, and are therefore not routinely encrypted (though they may or may not be authorized separately from control channels).
- *...... more details in the final paper ......*

(See Figure)

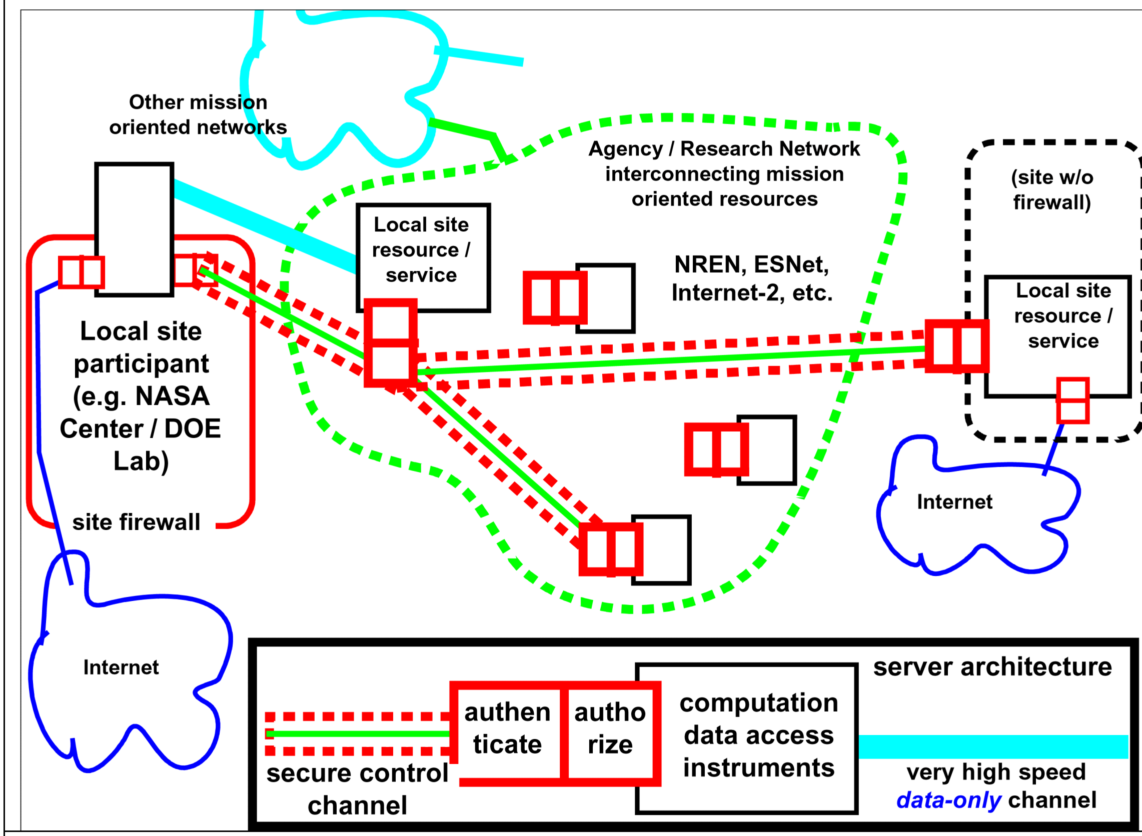

Figure 4. **Some Aspects of a Grid Security Model**

10 References and Acronyms

[1] Foster, I., and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*, edited by Ian Foster and Carl Kesselman. Morgan Kaufmann, Pub. August 1998. ISBN 1-55860-475-8. http://www.mkp.com/books_catalog/1-55860-475-8.asp

[2] Numerical Propulsion System Simulation (NPSS) - see http://hpcc.lerc.nasa.gov/npssintro.shtml

[3] For general information, see the security section in "**Bridging the Gap from Networking Technologies to Applications**." Workshop Co-sponsored by HPNAT & NRT (High Performance Network Applications Team & Networking Research Team of the Large Scale Networking (Next Generation Internet) Working Group). NASA Ames Research Center,

Moffett Field, Mountain View CA. Moffett Training and Conference Center, August 10 - 11, 1999. Published at http://www.nren.nasa.gov/workshop_home.html ("HPNAT/NRT Workshop")

[4] "**Real-Time Generation and Cataloguing of Large Data-Objects in Widely Distributed Environments**," W. Johnston, Jin G., C. Larsen, J. Lee, G. Hoo, M. Thompson, and B. Tierney (LBNL) and J. Terdiman (Kaiser Permanente Division of Research). Invited paper, International Journal of Digital Libraries - Special Issue on "Digital Libraries in Medicine". May, 1998. http://www-itg.lbl.gov/WALDO/

[5] Tierney, B. Lee, J., Crowley, B., Holding, M., Hylton, J., Drake, F., "**A Network-Aware Distributed Storage Cache for Data Intensive Environments**", Proceeding of IEEE High Performance Distributed Computing conference (HPDC-8), August 1999.

[6] MAGIC: "**The MAGIC Gigabit Network**." See: http://www.magic.net

[7] **TerraVision-2**: VRML based data fusion and browsing. (MAGIC consortium, NCAR, and NAVO: http://www.ai.sri.com/TerraVision/)

[8] **Globus** is a middleware system that provides a suite of services designed to support high performance, distributed applications. Globus provides:
- Resource Management: Components that provide standardized interfaces to various local resource management systems (GRAM) manage allocation of collections of resources (DUROC). All Globus resource management tools are tied together by a uniform resource specification language (RSL).
- Remote Access: Components that enable remote access to files (GASS and RIO) and executables (GEM).
- Security: Support for single sign-on, authentication, and authorization within the Globus system (GSI) and (experimentally) authorization (GAA).
- Fault Detection: Basic support for building fault detection and recovery into Globus applications.
- Information Infrastructure: Global access to information about the state and configuration of system components of an application (MDS).
- Grid programming services: Support writing parallel-distributed programs (MPICH-G), monitoring (HBM), etc.

ww.globus.org provides information about the Globus system.

[9] See "**Public Key Infrastructure Roadmap for the Department of Defense**." DoD Public Key Infrastructure Program Management Office. (http://www-pki.itsi.disa.mil/policy.html

[10] PKI: **Public-Key certificate Infrastructure**. Public-key cryptography involves two keys, whereby data encrypted with one key can only be decrypted with the other, and visa versa. In PKI one key (the public-key) is freely available and the other is kept private. In this way, material encrypted with the private key and decrypted with the public-key proves that the holder of the private key must have been the originator of the material. A certification authority generates a certificate containing the name (usually X.500 distinguished name) of

an entity (e.g. user) and that entity's public key. The CA then signs this "certificate" and publishes it (usually in an LDAP directory service). These are the basic components of PKI, and allow the entity to prove its identity, independent of location or system, by signing a token with the private key, handing the signed token to a system (e.g. as part of a login process), and then that system can verify the signer's identity by obtaining the identity certificate, extracting the entity's public key, and verifying the signature. The identity certificate (most commonly an X.509 certificate) is, in turn, verified by obtaining the CA's public key and using it to verify the contents. This later process is called digital signature and is accomplished by the certificate originator generating a unique hash of the certificate contents, and then encrypting that hash with the originator's private key. The hash is then appended to the certificate (or any other document) and may be used to both verify the originator's identity and the integrity of the contents (the hash function produces a "unique" hash for every byte string). For more information, see, e.g., RSA Lab's "**Frequently Asked Questions About Today's Cryptography**" http://www.rsa.com/rsalabs/faq/

[11] GSS: "**Generic Security Service Application Program Interface**", John Linn, Sep. 1993. Available at http://ds.internic.net/rfc/rfc1508.txt. Also see more recent and related drafts at the IETF Common Authentication Technology home page (http://www.ietf.cnri.reston.va.us/html.charters/cat-charter.html) and at http://www.ietf.cnri.reston.va.us/ids.by.wg/cat.html.

[12] GAA: "**Generic Authorization and Access control API**" (GAA API). IETF Draft. http://ghost.isi.edu/info/gss_api.html)

[13] **Akenti**: "**Certificate-based Access Control for Widely Distributed Resources**," Mary Thompson, William Johnston, Srilekha Mudumbai, Gary Hoo, Keith Jackson, Usenix Security Symposium '99. Mar. 16, 1999. (See http://www-itg.lbl.gov/Akenti)

[14] TLS: "The TLS Protocol, Version 1.0," RFC 2246, T. Dierks, C. Allen, January 1999.

[15] SSH: Tatu Ylonen. "**The SSH (Secure Shell) remote login protocol**". ftp://ietf.cnri.reston.va.us/internet-drafts/draft-ylonen-ssh-protocol-00.txt. See also "SSH (Secure Shell) Remote Login Program" (http://www.cs.hut.fi/ssh/).

[16] PKCS#12: " PKCS 12 v1.0: Personal Information Exchange Syntax," RSA Laboratories, June 1999. See also: http://www.rsasecurity.com/rsalabs/pkcs/pkcs-12/.